

Percepción de la adopción de la tecnología en los egresados de programas de humanidades de la Universidad de Caldas (Colombia) utilizando analítica predictiva en minería de datos

Perception of the adoption of technology in graduates of humanities programs of the University of Caldas (Colombia) using predictive analytics in data mining

Oscar Mauricio BEDOYA Herrera [1](#); Diego LOPEZ Franco [2](#); Carlos Eduardo MARULANDA Echeverry [3](#)

Recibido: 11/11/2017 • Aprobado: 30/11/2017

Contenido

- [1. Introducción](#)
- [2. Metodología](#)
- [3. Análisis de resultados](#)
- [4. Conclusiones](#)
- [Agradecimientos](#)
- [Bibliografía](#)

RESUMEN:

El presente artículo presenta los resultados del uso de técnicas de predicción usando el paquete de minería de datos DBMS_PREDICTIVE_ANALITICS basado en correlaciones multivariadas sobre el conjunto de datos de los egresados de programas de humanidades de la Universidad de Caldas (Colombia) entre los años 2010 y 2015 con las características socio educativas de los mismos. Se destaca que hay deficiencia en la adopción de la tecnología de los egresados.

Palabras clave: Minería de datos, técnicas de clasificación, egresados, súper modelos.

ABSTRACT:

This article presents the results of the use of prediction techniques using the data mining package DBMS_PREDICTIVE_ANALITICS based on multivariate correlations on the data set of graduates of humanities programs of the University of Caldas (Colombia) between 2010 and 2015 with the socio-educational characteristics of the same. It is highlighted that there is a deficiency in the adoption of the technology of the graduates.

Keywords: Data mining, classification techniques, graduates, super models.

1. Introducción

La cantidad de datos almacenados en diversos medios computacionales generados en las entidades educativas han aumentado considerablemente, es así como se ha generado la necesidad de aplicar metodologías y técnicas para el análisis de datos cuyo objetivo sea encontrar relaciones ocultas que permitan descubrir información y conocimiento.

En este orden de ideas es necesario que las investigaciones en el sector de la educación se encaminen en la aplicación de nuevos modelos de Minería de Datos que permitan el aprovechamiento de la información (Romero, y otros, 2008).

Para el caso de la educación universitaria y específicamente los egresados de programas humanísticos de la Universidad de Caldas en Manizales Colombia, es esencial descubrir cuáles son las variables académicas, socioeducativas y/o personales que influyen en las diversas percepciones existentes ante el enorme cambio que supone la incorporación y adopción de las tecnologías de la información y la comunicación (TIC) en el ámbito profesional y que condicionan directa o indirectamente su proyección futura. Las organizaciones educativas deben establecer herramientas que les ayuden en la toma de decisiones (Guitart Hormig, 2014) anticipándose a los problemas futuros a partir de la información obtenida de la minería de datos.

Las técnicas de minería de datos permiten indagar, explorar y analizar los datos provenientes de diversas fuentes. Según (Pérez-Palacios et al., 2014), la minería de datos es un factor determinante en el proceso conocido como generación de conocimiento en bases de datos. En esta investigación se utilizó el repositorio histórico de datos de la primera encuesta de egresados, la encuesta vía electrónica luego de tres años de graduado y las variables tanto de rendimiento académicas como sociales registradas cuando un determinado egresado era estudiante.

1.1. Minería de datos

Según (Hoffer, 2005) la minería de datos es la integración de todas las bases de datos soportadas por tecnologías emergentes de open data (datos abiertos), big data (Volumen alto de datos) y visualización computacional.

Representan avanzadas técnicas que permiten detectar relaciones entre los datos y obtener información no evidente: patrones, predicciones y asociaciones (Gómez y Suarez, 2010).

La minería de datos se considera como el proceso de seleccionar, explorar, modificar y valorar grandes cantidades de datos con el objetivo de descubrir conocimiento (Pérez y Santin, 2006). (Rosado y Rico, 2010) proponen que el proceso de explotación a través de la minería debe ser automático o semiautomático. Los modelos hallados deben ser significativos demostrando cierto patrón o regla de comportamiento. Las aplicaciones más utilizadas son las que necesitan algún tipo de predicción.

Según (PhridviRaj y GuruRao, 2013) la minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos. La generación de un modelo de minería de datos forma parte de un proceso mayor que incluye desde la formulación de preguntas acerca de los datos y la creación de un modelo para responderlas, hasta la implementación del modelo en un entorno de trabajo.

Complementa (Tsai, 2013), explicando que la minería de datos es un campo interdisciplinario que combina la inteligencia artificial, la gestión de bases de datos, la visualización de datos, el aprendizaje automático y algoritmos matemáticos y estadísticos. El conocimiento oculto se

genera a partir de un historial de datos que pueden modelar los aspectos pasados para predecir comportamientos y acciones futuras. Los datos deben estar debidamente clasificados, para encontrar las respectivas reglas y claves que esconden (Denesuk et al., 2003).

Según (Natek & Zwillig, 2014), los pasos para el análisis de la minería de datos, contemplan inicialmente la creación de los datos conjuntos, como segundo paso el definir la herramienta de minería de datos a utilizar, luego evaluar las técnicas a utilizar y finalmente, analizar los datos por cada modelo.

Es importante entender que el elemento clave para que el proceso de descubrimiento de conocimiento tenga éxito, no es solamente contar con datos almacenados sino como lo indica (Conesa & Curto 2010) es saber analizar y organizar los datos correctamente.

1.2. Herramientas de minería de datos

Las categorías básicas de las técnicas de minería de datos se pueden reunir en: clasificación, asociación, secuenciación y clusterización.

Para (Tumero, 2011) la clasificación incluye los procesos de minería de datos que buscan reglas para definir si un ítem o un evento pertenecen a un subconjunto particular o a una clase de datos. Esta técnica, probablemente la más utilizada, incluye dos subprocesos: la construcción de un modelo y la predicción.

Enuncia (Tumero, 2011) que en términos generales, los métodos de clasificación desarrollan un modelo compuesto por reglas IF-THEN (si-entonces) y se aplican perfectamente, por ejemplo, para encontrar patrones de compra en las bases de datos de los clientes y construir mapas que vinculan los atributos de los clientes con los productos comprados. Con un conjunto apropiado de atributos predictivos, el modelo puede identificar los clientes con mayor propensión a realizar una determinada compra durante el próximo mes. Un caso típico de clasificación es el de dividir una base de datos de compañías en grupos homogéneos respecto a variables como "posibilidades de crédito" con valores tales como "bueno" y "malo".

(Westphal & Blaxton, 1998) especifican que este tipo de herramientas son utilizadas en los paquetes de minería de datos donde se emplean distintos parámetros estadísticos; a continuación se mencionan de manera global algunos de ellos: medias, varianzas, y correlaciones que forman un conjunto de métodos estadísticos y matemáticos, destinados a describir e interpretar los datos que provienen de la observación de varias variables, estudiadas conjuntamente.

El análisis de la varianza contrasta si existen diferencias significativas entre las medidas de una o más variables continuas en grupo de población distintos.

La regresión define la relación entre una o más variables y un conjunto de variables que son predictores de las primeras, aproximando y pronosticando valores continuos.

El análisis clúster permite clasificar una población en un número determinado de grupos, en base a semejanzas y desemejanzas de perfiles existentes entre los diferentes componentes de dicha población.

El análisis discriminante, método de clasificación de individuos en grupos que previamente se han establecido, y que permite encontrar la regla de clasificación de los elementos de estos grupos, y por tanto identificar cuáles son las variables que mejor definan la pertenencia al grupo.

(Espinosa, 2010) habla de los métodos basados en árboles de decisión, el método Chaid (Chi Squared Automatic Interaction Detector) es un análisis que genera un árbol de decisión para predecir el comportamiento de una variable, a partir de una o más variables predictoras, de forma que los conjuntos de una misma rama y un mismo nivel son disjuntos. Es útil en aquellas situaciones en las que el objetivo es dividir una población en distintos segmentos basándose en algún criterio de decisión.

La lógica difusa es una generalización del concepto de estadística. La estadística clásica se basa en la teoría de probabilidades, a su vez ésta en la técnica conjuntista, en la que la relación de pertenencia a un conjunto es dicotómica (el 2 es par o no lo es).

La serie temporal, es el conocimiento de una variable a través del tiempo para, a partir de ese conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales, poder realizar predicciones. Suelen basarse en un estudio de la serie en ciclos, tendencias y estacionalidades, que se diferencian por el ámbito de tiempo abarcado, para por composición obtener la serie original. Se pueden aplicar enfoques híbridos con los métodos anteriores, en los que la serie se puede explicar no sólo en función del tiempo sino como combinación de otras variables de entorno más estables y, por lo tanto, más fácilmente predecibles.

Las técnicas bayesianas, se basan en probabilidades condicionales. Utiliza el teorema de Bayes, como una fórmula que calcula una probabilidad por contar la frecuencia de los valores y combinaciones de valores en los datos históricos. El Test de hipótesis o contraste de hipótesis se define como el procedimiento estadístico mediante el cual se investiga la verdad o falsedad de una hipótesis acerca de una población o poblaciones.

El análisis multivariante según (Salvador, 2000) es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariados en el sentido de relacionar varias variables medidas para cada individuo u objeto estudiado. Su razón de ser radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo información que los métodos estadísticos univariantes y bivariantes son incapaces de conseguir.

El análisis clúster se centra en la agrupación de datos para poder llevar a cabo segmentación de consolidados basados en la medición. Permiten llevar a cabo la identificación de grupos de datos con características y patrones similares.

La asociación de reglas busca elementos que tienden a ocurrir en los datos y especifica las reglas que rigen su ocurrencia. Los modelos lineales generalizados (GLM) orientados a la clasificación y la regresión, constituye la regresión logística para la clasificación de objetivos binarios y de regresión lineal para los objetivos continuos. Es compatible con los límites de confianza para las probabilidades de predicción.

2. Metodología

A continuación se relacionan los aspectos metodológicos utilizados para el desarrollo de la investigación que soportan los resultados presentados en este artículo.

2.1. Descripción de la base de datos

La población de datos analizada consta de 2694 registros correspondientes a los alumnos egresados de programas de humanidades: (Antropología, Artes Plásticas, Derecho, Desarrollo Familiar, Licenciaturas Artes Escénicas, Licenciatura en Ciencias Sociales, Trabajo Social, Sociología entre otras) de la Universidad de Caldas entre en los años 2010 y 2015 como se muestra en la Figura 1.

Figura 1
Muestreo de datos.

PROGRAMA	EGRESADOS
ANTROPOLOGIA	154
ARTES PLASTICAS	88
DERECHO	419
DESARROLLO FAMILIAR	258
LICENCIATURA EN ARTES ESCENICAS CON ENFASIS EN TEATRO	74
LICENCIATURA EN CIENCIAS SOCIALES	233
LICENCIATURA EN EDUCACION AMBIENTAL	53
LICENCIATURA EN EDUCACION BASICA CON ENFASIS EN EDUCACION FISICA	332
LICENCIATURA EN EDUCACION FISICA Y RECREACION	29
LICENCIATURA EN FILOSOFIA Y LETRAS	20
LICENCIATURA EN LENGUAS MODERNAS	303
LICENCIATURA EN MUSICA	107
SOCIOLOGIA	116
TRABAJO SOCIAL	508

Fuente: los autores

2.2. Muestreo

Se alcanzaron los repositorios de datos internos y externos; la fuente interna se seleccionó de la base de datos histórica de los egresados de la Universidad de Caldas en los programas de Humanidades en los últimos cinco años, adicionalmente como fuente externa se condensó la encuesta de seguimiento a momento cero a graduados que tiene como fin recopilar y analizar el desarrollo profesional y personal de los recién profesionales de educación superior frente a las áreas de conocimiento adquiridos en su trayecto de formación universitaria así como de bienestar vividos en su actividad como estudiantes.

De la generalidad de variables contenidas en las encuestas digitales (Munuera P., & Roesler, J. (2007) se consideraron para el análisis aquellas que presentaban mayor particularidad en la validez de los datos, tal como se observa en la tabla 1.

Tabla 1
Estructura de la base de datos

Atributo	Descripción
ID	Identificador único del egresado
ESTADOCIVIL	Especifica el estado civil del egresado, el cual puede ser: soltero, casado o en unión libre
GENERO	Especifica si el egresado es masculino o femenino
RANGOEDAD	Especifica el rango de edad del egresado al momento de graduarse: Menor de 18 años, entre 18 y 22 años, entre 23 y 25 años, mayor de 26 años.
NIVELPADRE	Especifica el nivel educativo del padre del egresado, el cual puede ser: nunca estudió, primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, educación técnica, educación universitaria incompleta, educación universitaria completa o educación de postgrado.

NIVELMADRE	Ídem al nivelpadre
TIEMPOEMPLO	Especifica el tiempo que el egresado uso para para conseguir su primer empleo:menos de tres meses, entre tres y seis meses, más de seis y hasta 1 año o más de 1 Año
ETNIA	Especifica la raza originaria del egresado: mestizo, indio o afrocolombiano
UTILESHABI	Determina si las habilidades adquiridas en el pregrado son valederas en su actividad profesional desde: muy útiles, útiles, poco útiles o nada útiles
PERFORMANCE	Indicador que establece el rendimiento del egresado a través de su historia como estudiante de pregrado. Valores Rendimiento Regular o Bajo, Rendimiento Aceptable, Rendimiento Bueno, Rendimiento Sobresaliente
USOTECNOLO	Especifica la caracterización del uso de la tecnología que el egresado asume en cuando sale del pregrado. Categorizaciones:(Muy satisfecho, Insatisfecho, satisfecho, muy insatisfecho)
GESTIONAR	Atributo de valoración la que el egresado percibe sus capacidades de enfrentar la administración de operaciones en su rol específico. Los atributos son los mismos que el uso de la tecnología.
INNOVAR	Atributo de valoración la que el egresado percibe sus capacidades de generar nuevas herramientas, ideas, posturas en su rol específico. Los atributos son los mismos que el uso de la tecnología e innovar.
RESPROBLEM	Atributo de valoración la que el egresado percibe sus capacidades de resolver y enfrentarse a problemas de diversa índole. Categorizaciones: (Muy satisfecho, Insatisfecho, satisfecho, muy insatisfecho)

Fuente: los autores

2.3. Herramienta de análisis predictivo de los datos

Lo que se quiere descubrir: son los atributos socioculturales y/o de rendimiento académico más influyentes que pueden condicionar en la valoración de la adopción de la tecnología para los recién egresados de ciencias humanísticas de la Universidad de Caldas.

Para el análisis predictivo se utilizó el DBMS_PREDICTIVE_ANALYTICS en lenguaje PL/SQL entorno Oracle. Esta herramienta de minería tiene como meta encontrar patrones en los datos y las relaciones en la bodega de información cuyo objetivo será brindar pautas que puedan utilizarse para predecir eventos futuros.

El paquete viene con las siguientes funciones: (EXPLAIN y PREDICT). EXPLAIN crea un modelo de importancia de los atributos, analiza el conjunto de datos para determinar el valor explicativo de cada atributo. Cuanto mayor sea el valor explicativo para un atributo, más fuerte es la relación entre el mismo y el resultado. Seguidamente se devuelven una lista de atributos clasificados en orden relativo de su impacto en la predicción. Esta información se deriva de los datos del modelo en beneficio del patrón de atributo de importancia

La función de la minería de datos de análisis predictivo analiza los datos, encuentra patrones, y predice los resultados con un nivel de confianza asociado. La confianza es importante, ya que pondera la certeza de las predicciones. El rango de valores es de 0 a1. Cuanto mayor sea el

valor de confianza, más segura será la predicción. El proceso de ejecución del paquete predictivo mostrado en la figura 2 corresponde a la relación ordenada de atributos que implican mayor correlación en el descubrimiento de la información frente al atributo (ADOPTAR_TECNOLOGIA). Los resultados se pueden observar en la figura 2.

Figura 2

Salida de Correlación para el atributo Adoptar_Tecnologia

```

***
*** DBMS_PREDICTIVE_ANALYTICS.EXPLAIN
*** Print the ASSESSMENT table's columns
*** and their correlation to ADOPTAR TECNOLOGIA c
***

```

Procedimiento PL/SQL terminado correctamente.

ATTRIBUTE_NAME	ATTRIBUTE_SUBNAME	EXPLANATORY_VALUE	RANK
USOTECNOLO		.386921499	1
GESTIONAR		.291253474	2
INNOVAR		.264744571	3
RESPROBLEM		.245919376	4
ACTUALIZAR		.237743868	5
ANALISIS		.187715702	6
COMPRESIO		.17899393	7
ESCRITURA		.150006704	8
ACEPCULTUR		.148414529	9
COMORAL		.145030508	10
FORTALEZA		.067422189	11
DEBILIDAD		.03795099	12
FACULTAD		.007737838	13
RAZONNOU		.007470078	14
RAZONUOLU		.0058197	15
TIPOESTUD		.005212453	16
PROGRAMA		.004967995	17
PROYUIDA1		.001292267	18
FACTORING		.000052011	19
INGORAL		.000651209	20
TIEMBACPRE		.000595814	21
UTILICONOC		.000581122	22
PROYUIDA3		.000522699	23
ENTREVISTA		0	24
JINUNIEMP		0	24
PREFERPRIM		0	24
TERMINAB		0	24
ETNIA		0	24
OCUPAMADRE		0	24
NIVEDUCMAD		0	24

Fuente: los autores

Los resultados muestran que los atributos (USOTECNOLO), (GESTIONAR), (INNOVAR) Y (RESPROBLEM) son los mejores predictores de (ADOPTAR_TECNOLOGIA).

Los tres primeros campos de la columna (RANK) que tienen un (EXPLANATORY_VALUE) positivo más cercano a uno, son los atributos que se pueden utilizar para predecir la percepción de un individuo frente la adopción de la tecnología en su futuro profesional.

La percepción de un egresado en humanidades frente a la adopción de la tecnología está estrechamente demarcado por el atributo correspondiente al uso de la misma (USOTECNOLO), seguido por las capacidades de enfrentar la administración de operaciones en su rol específico (GESTIONAR). La significancia del valor en el ranking originado por el EXPLANATORY_VALUE puede determinar valores cercanos de predicción y comportamientos según las correlaciones identificadas.

3. Análisis de resultados

El procedimiento PREDICT es usado para predecir los valores de una columna específica. La

entrada consiste en unos parámetros y una columna de destino, la columna que contiene los valores objetivo de predecir. Los datos de entrada deben contener algunos casos en los que se conoce el valor objetivo (es decir, no es nulo o indeterminado).

Los casos en que se conocen los valores de destino se utilizan para entrenar y probar el modelo. PREDICT retornara un valor previsto para cada caso, incluyendo aquellos en los que se conoce el valor.

La finalidad será producir predicciones para objetivos desconocidos. Los casos donde el objetivo es indeterminado, es decir, cuando el valor es nulo, no se tendrán en cuenta durante el entrenamiento del modelo.

Por tanto, el siguiente proceso se enfocara en determinar el grado en que influye el atributo (GESTION) con el factor de aprobación de adopción de la tecnología (ADOPTAR_TECNOLOGIA) utilizando un clasificador basado en el algoritmo Naïve Bayes.

La predicción de resultados se realizara para aquellos egresados cuyo valor en el atributo (GESTIONAR) sea "Insatisfecho" o "Muy Insatisfecho".

El procedimiento produjo una salida (tabla de parámetros result_table_name) y con tres atributos resultantes mostrados en la figura 3.

CASE_ID =Cada uno de los casos identificados en la columna.

PREDICTION =Valor predicho del atributo destino.

PROBABILITY =Probabilidad que la predicción sea correcta valor.

Figura 3

Valor de confianza de predicción del modelo.

```
SQL Plus
***
*** DBMS_PREDICTIVE_ANALYTICS.EXPLAIN
*** Print the ASSESSMENT table
*** PREDICT FOR ADOPTARTEC
***
*** Accuracy ***
.679707556
Procedimiento PL/SQL terminado correctamente.
REGISTROS_TOTALES
-----
                2694
1 fila seleccionada.
PREDICTION                                REGISTROS
-----
Muy Insatisfecho                            103
Insatisfecho                                501
Muy satisfecho                               698
Satisfecho                                  1392
```

Fuente: los autores

La tasa de valoración de la variable "Accuracy" mostrada en la figura 3 es de 67,97% muy representativa sobre el conjunto de clasificación. La predicción conduce entonces afirmar que para los 2694 egresados cuyo dato almacenado en el atributo (GESTIONAR) este determinado por la no aprobación del mismo predecirán una percepción de insatisfacción general de la percepción de la adopción de la tecnología (ADOPTAR_TECNOLOGIA) de "Muy Insatisfecho o Insatisfecho", en 604 de ellos.

Ahora bien, se realiza una consulta sobre las predicciones y los resultados del procedimiento de PREDICT para 17 estudiantes como muestreo de prueba ordenados por la probabilidad mostrados en figura 4:

Figura 4
Muestreo de predicción.

IDCEDUL	PREDICTION	PROBABILITY
4539442	Muy satisfecho	7,866E-001
4472865	Satisfecho	7,566E-001
4438602	Satisfecho	7,535E-001
4803384	Muy satisfecho	7,514E-001
4539365	Satisfecho	7,462E-001
4439485	Muy satisfecho	7,453E-001
4472047	Satisfecho	7,406E-001
454022	Satisfecho	7,309E-001
4547869	Muy satisfecho	7,25E-001
4339230	Satisfecho	7,22E-001
138631	Satisfecho	7,18E-001
4540022	Satisfecho	6,904E-001
1025898	Satisfecho	6,859E-001
3497408	Satisfecho	6,83E-001
6240126	Muy satisfecho	6,817E-001
4472247	Insatisfecho	5,698E-001
2438455	Insatisfecho	4,757E-001

17 filas seleccionadas.

Fuente: los autores

En la figura 4 se puede apreciar que la probabilidad de predicción oscila entre un 47% y un 78%, valores determinantes.

Por lo tanto, el egresado cuyo identificador 4472247, que tiene un rendimiento de 3 entre 1 y 5 siendo este último el rendimiento ideal, con un nivel de educación de su madre correspondiente a "Primaria Completa" y un nivel de educación de su padre de "Educación Técnica", que considera "Muy Útiles" los conocimientos adquiridos en la universidad y cuya percepción de gestionar es "Muy Insatisfecho", se proyecta en responder que su percepción de la adopción de la tecnología es "Insatisfecho", con un 56,98% por ciento de certeza.

Finalmente, se contrasta el valor objetivo original con el valor de predicción como se muestra en la figura 5 para un número determinado de registros en este caso 40 egresados. Los resultados que se observan en la figura 5.

Figura 5
Contraste de resultados valor objetivo vs valor de predicción

IDCEDUL	FACULTAD	PERCEPCION	PREDICCION	PROBABILITY
1053819279	TRABAJO SOCIAL	Muy satisf	Muy satisf	6,978E-001
1053821020	DERECHO	Muy satisf	Muy satisf	6,69E-001
1053822313	LICENCIATURA EN	Muy satisf	Muy satisf	7,092E-001
1053821139	LICENCIATURA EN	Satisfecho	Satisfecho	7,128E-001
1053821154	DESARROLLO FAMI	Satisfecho	Satisfecho	7,351E-001
1053824796	LICENCIATURA EN	Satisfecho	Satisfecho	6,596E-001
1053822295	LICENCIATURA EN	Insatisfec	Insatisfec	5,204E-001
1053824423	TRABAJO SOCIAL	Muy satisf	Muy satisf	7,3E-001
1144126614	DERECHO	Muy satisf	Muy satisf	7,16E-001
1144129689	DERECHO	Muy satisf	Muy satisf	7,122E-001
1125470690	SOCIOLOGIA	Satisfecho	Satisfecho	6,948E-001
1124853117	LICENCIATURA EN	Muy satisf	Muy satisf	7,603E-001
1053816833	DERECHO	Muy satisf	Muy satisf	6,255E-001
1061047052	LICENCIATURA EN	Muy satisf	Muy satisf	7,595E-001
1060650943	LICENCIATURA EN	Muy satisf	Muy satisf	7,591E-001
1060649297	LICENCIATURA EN	Satisfecho	Satisfecho	6,944E-001
1053812171	LICENCIATURA EN	Muy satisf	Muy satisf	7,351E-001
1053814166	TRABAJO SOCIAL	Satisfecho	Satisfecho	7,575E-001
1053815081	DERECHO	Insatisfec	Insatisfec	4,618E-001
1053817797	SOCIOLOGIA	Satisfecho	Satisfecho	3,011E-001
1053812497	ANTROPOLOGIA	Insatisfec	Insatisfec	3,321E-001
1053814609	DERECHO	Insatisfec	Insatisfec	6,661E-001
1053813437	TRABAJO SOCIAL	Muy satisf	Muy satisf	7,152E-001
1053813473	DERECHO	Insatisfec	Insatisfec	6,148E-001
1053813333	TRABAJO SOCIAL	Satisfecho	Satisfecho	7,534E-001
1053813335	DERECHO	Satisfecho	Satisfecho	7,069E-001
1053815241	DERECHO	Insatisfec	Insatisfec	3,314E-001
1,0538E+10	SOCIOLOGIA	Muy satisf	Muy satisf	7,745E-001
1053812982	SOCIOLOGIA	Satisfecho	Satisfecho	3,619E-001
1053812659	LICENCIATURA EN	Satisfecho	Satisfecho	3,404E-001
1053813830	LICENCIATURA EN	Satisfecho	Satisfecho	7,486E-001
1054988081	DERECHO	Insatisfec	Insatisfec	4,606E-001
1053814798	TRABAJO SOCIAL	Satisfecho	Satisfecho	7,397E-001
1054988351	LICENCIATURA EN	Satisfecho	Satisfecho	7,26E-001
1054988499	ANTROPOLOGIA	Satisfecho	Satisfecho	7,144E-001
1054987130	ARTES PLASTICAS	Satisfecho	Satisfecho	4,127E-001
1101447955	LICENCIATURA EN	Satisfecho	Satisfecho	7,296E-001
1101448189	LICENCIATURA EN	Muy satisf	Muy satisf	7,565E-001
1094900697	TRABAJO SOCIAL	Satisfecho	Satisfecho	7,23E-001

39 filas seleccionadas.

Fuente: los autores

Los resultados arrojados por el contraste de las muestras originadas en los registros de la base de datos determinan que los valores del atributo (ADOPTAR_TECNOLOGIA) correlacionados con el atributo (GESTIONAR) en contraste con los resultados del algoritmo de predicción tienen un éxito en el pronóstico efectivo del 68% para la totalidad de los registros estudiados.

Ahora bien se agrupan los resultados tal como se observa en la figura 6.

Figura 6
Agrupación de resultados de predicción por programa académico

FACULTAD	PROBABILIDAD	PREDICCIÓN
LICENCIATURA EN CIENCIAS SOCIALES	5,396E-001	Insatisfecho
ARTES PLASTICAS	4,839E-001	Insatisfecho
SOCIOLOGIA	4,642E-001	Insatisfecho
LICENCIATURA EN EDUCACION AMBIENTAL	5,393E-001	Insatisfecho
TRABAJO SOCIAL	5,05E-001	Insatisfecho
LICENCIATURA EN ARTES ESCENICAS CON ENFASIS E	5,572E-001	Insatisfecho
ANTROPOLOGIA	4,74E-001	Insatisfecho
LICENCIATURA EN LENGUAS MODERNAS	5,363E-001	Insatisfecho
LICENCIATURA EN EDUCACION FISICA Y RECREACION	5,823E-001	Insatisfecho
DERECHO	5,338E-001	Insatisfecho
LICENCIATURA EN FILOSOFIA Y LETRAS	5,626E-001	Insatisfecho
DESARROLLO FAMILIAR	5,252E-001	Insatisfecho
LICENCIATURA EN EDUCACION BASICA CON ENFASIS	4,981E-001	Insatisfecho
LICENCIATURA EN MUSICA	5,662E-001	Insatisfecho

14 filas seleccionadas.

Fuente: los autores

La figura 6 muestra:

El 53.96% de los recién egresados del programa en "LICENCIATURA EN CIENCIAS SOCIALES" son predecibles como "Insatisfecho" en la adopción de la tecnología conforme a su deficiente capacidad gestionar.

El 50.5% de los recién egresados del programa en "TRABAJO SOCIAL" son predecibles como "Insatisfecho" en la adopción de la tecnología conforme a su deficiente capacidad gestionar

El 53.33% de los recién egresados de la facultad de "DERECHO" son predecibles como "Insatisfechos" con la adopción de la tecnología conforme a su deficiente capacidad gestionar.

4. Conclusiones

Para hacer minería y entender lo que no es evidente se requiere de una precisa transformación de datos. La preparación de estos en el tradicional esquema de descubrimiento de conocimiento es considerado ETL puro. De hecho, en función de los algoritmos que se usan en minería estos pueden variar en complejidad y necesitar de ajustes que ocasionan un arduo y duro trabajo por parte de los investigadores. Es aquí donde entra el concepto y aprovechamiento de un modelo de orden superior basado en los motores de bases de datos actuales más importante del mercado y que son herramienta fundamental en la producción de la minería de datos como lo es en este caso el paquete DBMS_PREDICTIVE_ANALYTICS. Esta activa herramienta realiza la preparación automática de datos, mientras que el modelo está en construcción. Por consiguiente añade más funcionalidad a la creación del simple modelo y da una experiencia superior fuera de confort del usuario.

Sorprende el tiempo de proceso de los datos, en total menos de 20 segundos duraron los procesos sin importar el volumen considerable de los registros, factor fundamental de aprovechamiento para futuros trabajos que ameriten la integración de nuevos conceptos como los servicios de la computación en la nube y la inteligencia de negocios.

La minería de datos hace parte de la tendencia en la inteligencia de negocios y particularmente sorprende con herramientas de alto impacto en los objetos de predicción (Vieira; et al., 2009), elemento fundamental por el que las entidades mundiales tienen interés en el desarrollo de soluciones.

Agradecimientos

Los resultados presentados en el artículo hacen parte del proyecto de investigación titulado: Explotación de datos masivos en la bodega de datos de egresados de la Universidad de Caldas y su integración con procesos de Gestión del Conocimiento e Inteligencia de Negocios.

Bibliografía

- Denesuk M., Frederick D., Snow K., Rajagopalan S., Tomkins A. (2003). *Knowledge based data mining system*. Patente de Patent No. US 2003/0212675 A1 (USA) Patents Services. (11/10/2012). 11 p.
- Espinosa R. (2010). DataMining o Minería de Datos. *Dataprix*.
- Gómez A., Suarez C., eds. *Herramientas Practicas para la gestión*. 2010.
- Guitart Hormigo I., & Conesa I Caralt, J. (2014). Uso de analítica para dar soporte a la toma de decisiones docentes. Jornadas de Enseñanza Universitaria de la Informática.
- Hoffer J., Prescott, M., & McFadden, F. (2005). *Modern Database Management*. Pearson.
- Conesa J., & Curto, J. (2010) Introducción al Business Intelligence. *Editorial UOC*.
- Munuera P., & Roesler, J. (2007). Las Tecnologías de Información y las comunicaciones (TICs) en la enseñanza de Trabajo Social y Servicios Sociales desde la perspectiva de la Convergencia Europea de los Estudios Superiores (CEES). En Fernán-dez-Valmayor Crespo, A.; Sanz Cabrerizo, A.; y Merino Granizo, J.: IV Jornada. CV-UCM. Ex-periencias en el CV: resultados. Madrid, Editorial Complutense, pp. 55-62.
- Natek, S., & Zwilling, M. (2014). Student data mining solution–knowledge management system related to higher education institutions. *Expert Systems with Applications*(41), 6400–6407.
- Pérez-Palacios, T., Caballero, D., Caro, A., Rodríguez, P., & Antequera, T. (2014). Applying data mining and Computer Vision Techniques to MRI to estimate quality traits in Iberian hams. *Journal of Food Engineering*(131), 82–88.
- Pérez C., Santin D. (2006). *Data Mining Soluciones con Enterprise Miner*.
- PhridviRaj M.S.B., GuruRao C.V. (2013). Data mining past, present and future a typical survey on data streams. *Procedia Technology*, 12: 255 – 263.
- Romero, C., Ventura, S. and Garcia, E., "Data mining in course management systems: Moodle case study and Tutorial". *Computers & Education*, Vol. 51, No. 1. pp.368- 384. 2008
- Rosado A., Rico D. (2010). Inteligencia de negocios. *Scientia Et Technica* 2010 XVI(44).
- Salvador M. (2000). Introducción al Análisis Multivariante. *Estadística en línea*.
- Tsai H. (2013). Knowledge management vs. data mining: Research trend, forecast and citation approach. *Expert Systems with Applications*, 40: 3160-3173.
- Tumero I. (2011). El arte de sacar conocimiento de grandes volúmenes de datos. *Minería de Datos*. pp. 19 – 45.
- Vieira; L., Ortiz; L., Ramirez S. (2009). Introducción a la Minería de Datos. *Rio de Janeiro: E-Papers Servicios Editoriales*.
- Westphal C., Blaxton T. (1998). *Data mining solutions*, in *Jhon Wiley & Sons*.

-
1. MsC en Ingeniería Computacional, Universidad de Caldas. Profesor de la Facultad de Ingeniería de la Universidad de Caldas, Manizales, Caldas, Colombia. oscar.bedoya@ucaldas.edu.co
 2. MsC en Educación, Universidad Católica de Manizales. Profesor de la Facultad de Ingeniería de la Universidad de Caldas, Manizales, Caldas, Colombia. diego.lopez_f@ucaldas.edu.co
 3. PhD. Ingeniería-Industria y Organizaciones, Universidad Nacional de Colombia sede Manizales. Profesor Facultad de Administración de la Universidad Nacional de Colombia sede Manizales y de la Facultad de Ingeniería de la Universidad de Caldas, Manizales, Caldas, Colombia. cemarulandae@unal.edu.co y carloose@ucaldas.edu.co
-

[Índice]

[En caso de encontrar algún error en este website favor enviar email a [webmaster](#)]

©2018. revistaESPACIOS.com • Derechos Reservados